Upper bounds on the rate of linear q-ary k-hash codes

Stefano Della Fiore
Department of Computer Science
University of Salerno
sdellafiore@unisa.it

Marco Dalai
Department of Information Engineering
University of Brescia
marco.dalai@unibs.it

***Abstract*—This paper presents new upper bounds on the rate of linear k-hash codes in $\mathbb{F}_q^n$, $q \geq k$, that is, codes with the property that any k distinct codewords are all simultaneously distinct in at least one coordinate.**

I. Introduction

A q-ary code C of length n is a subset of $\{0, 1, \ldots, q-1\}^n$. We denote the rate of a q-ary code C of length n as

$$R = \frac{1}{n} \log_q |C| \,.$$

Let $q \geq k \geq 3$ and $n \geq 1$ be integers, and let C be a q-ary code of length n with the property that for any k distinct elements (codewords) we can find a coordinate in which they all differ. A subset C with this property is called (q, k)-hash code of length n. In particular, $(q, 3)$-hash codes are known as q-ary trifferent codes (or just trifferent codes when $q = 3$). The problem of finding upper and lower bounds for the maximum size of (q, k)-hash codes is a fundamental problem in theoretical computer science and information theory. It appears, as the name suggests, in the study of families of perfect hash functions and in the study of the zero-error capacity of some discrete channels with list decoding, see [1], [2], [3], [4], [5] for more details (see also [6] for a related problem).

An elementary double counting argument, as shown in [2], gives the following bound on the cardinality of (q, k)-hash codes:

$$|C| \leq (k-1) \left(\frac{q}{k-1} \right)^n \text{ for every } q \geq k \geq 3 \,. \tag{1}$$

In 1984 Fredman and Komlós [1] improved the bound in (1) for every $q = k \geq 4$ and sufficiently large n, obtaining the following result:

$$|C| \leq \left((q-k+2)^{q^{\underline{k-1}}/q^{k-1}} \right)^{n+o(n)} , \tag{2}$$

where $q^{\underline{k-1}} = q(q-1)\cdots(q-k+2)$. Fredman and Komlós also provided, using standard probabilistic methods, the following lower bound:

$$|C| \geq \left(\left(1 - \frac{q^{\underline{k}}}{q^k} \right)^{-1/(k-1)} \right)^{n+o(n)} . \tag{3}$$

A generalization of the upper bounds given in equations (1) and (2) was derived by Körner and Marton [2] in the form

$$|C| \leq \min_{0 \leq j \leq k-2} \left(\left(\frac{q-j}{k-j-1} \right)^{q^{\underline{j+1}}/q^{j+1}} \right)^{n+o(n)} . \tag{4}$$

In 1998, Blackburn and Wild [4] (see also [7]) improved the bound of Körner and Marton for every q sufficiently larger than k, proving

$$|C| \leq (k-1) q^{\lceil \frac{n}{k-1} \rceil} \,. \tag{5}$$

Much effort has been spent during the years to refine the bounds given in (1), (2) and (4). See for example [8], [9] and the recent breakthrough [10] for case of $q = k = 3$, [11], [3], [12] in case of $q = k = 4$, [13] in case of $q = k = 5, 6$, and [2], [14], [15], [5] for $q \geq k \geq 5$. However, to the best of our knowledge, for q sufficiently larger than k no improvements over the upper bound (5) have been obtained.

For the sake of completeness, it is worth noting that some improvements on the lower bound given in (3) on the largest size of (q, k)-hash codes have been recently obtained in [16] for $q \in [4, 15]$, all integers between 17 and 25, and for a sufficiently large q. For $q = k = 3$ the best known lower bound is due to Körner and Marton [2] and it is equal to $(9/5)^{n/4+o(1)}$.

In contrast, no exponential improvement has been made on the simple bound given in (1) for $q = k = 3$ (see for example [17] for a discussion on the intractability of this problem even with some recent powerful techniques such as the slice-rank method). Until recently, only improvements on the multiplicative constant had been obtained (see [8], [9]), while a polynomial improvement has been obtained in a beautiful recent work by Bhandhari and Kheta [10].

However, if we restrict the codes to be linear, i.e., we require C to be a linear subspace of $\mathbb{F}_3^n$, exponential improvements have been recently obtained. Upper bounds on the rate of linear trifferent codes have been considered first in [18], where it was proved that for some $\epsilon > 0$,

$$\begin{aligned} |C| &\leq 3^{(\frac{1}{4}-\epsilon)n} \\ &\approx 1.3161^n \,. \end{aligned} \tag{6}$$

2024 IEEE International Symposium on Information Theory (ISIT) | 979-8-3503-8284-6/24/$31.00 ©2024 IEEE | DOI: 10.1109/ISIT57864.2024.10619541

This result was then improved in [19], where connections with minimal codes were used to show that

$$|C| \leq 3^{n/4.5516+o(n)} \approx 1.2731^n, \tag{7}$$

where the constant 1.2731 is the numerical solution of an equation that we will explain in Section II. The authors also showed that there exist linear trifferent codes of length n and size $\frac{1}{3}(9/5)^{n/4}$, matching, asymptotically in n, the best known lower bound on trifferent codes (without the linearity constraint) obtained in [2].

We note that when q is small compared to $k > 3$, no linear k-hash codes of dimension 2 exist. In particular, Blackburn and Wild [4] showed that this is true for every $q \leq 2k-4$ (so, in particular, the case $q = k > 3$ is of no interest). The authors in [20] improved this result for $k \geq 9$ showing that when q is a square, $\sqrt{q} > 5$, no linear k-hash codes of dimension 2 exist whenever $q \leq \left(\frac{k-1}{2}\right)^2$. Hence in these regimes, we know that linear k-hash codes in $\mathbb{F}_q^n$ are relatively *simple objects* since their asymptotic rates are equal to zero.

In this paper we provide upper bounds on the rate of linear k-hash codes in $\mathbb{F}_q^n$ for general values of $q \geq k \geq 3$. These are the first known (non-trivial) such bounds. For $q = k = 3$ they recover the best known result of [19] given in equation (7). Also, in the range of q much larger than k, they improve the general bound of equation (5) (in terms of code rate as $n \to \infty$).

II. A simpler proof for $q = k = 3$

In this section, we present a re-derivation of (6) and (7) by a straightforward application of a method already presented[1] in [21]. This simpler approach is the starting point for the extension to the general case $q \geq k \geq 3$ which is then presented in the next section.

The idea is to modify slightly the proof of [21, Corollary 2.1]. The main tool to be used is Jamison's bound [22].

Lemma 1 ([22]): Let $q \geq 3$ be a prime power, and let $\mathcal{H}$ be a set of hyperplanes in $\mathbb{F}_q^m$ whose union is $\mathbb{F}_q^m \setminus \{0\}$. Then $|\mathcal{H}| \geq (q-1)m$.

Let C be a linear trifferent code of dimension m and length n. Let G be the $m \times n$ generator matrix, let d be the minimum Hamming distance of the code, and let x be a codeword of weight d. Finally call $u \in \mathbb{F}_3^m$ the information vector associated to x, that is, assume $x = uG$, and also assume without loss of generality (by appropriate sorting and re-scaling of the columns of G) that x has 0s in the last $n-d$ coordinates and 1s in the first d coordinates. Then, since the code is trifferent, any codeword different from 0 and x must have a coordinate equal to 2 among the first d ones. So, if we call g_i the i-th column of G, the d affine subspaces defined by

$$H_i = \{v \in \mathbb{F}_3^m : v \cdot g_i = 2\}, i = 1, \ldots, d$$

cover the set $\mathbb{F}_3^m$ with the exception of 0 and u. Adding another subspace $H_{d+1} = \{v \in \mathbb{F}_3^m : v \cdot g_1 = 1\}$ we also covers u, still leaving out 0. By Jamison's bound, $d+1 \geq 2m$. In terms of rates and relative minimum distance $\delta = d/n$ this becomes

$$R \leq \frac{1}{2}\delta + o(1). \tag{8}$$

The rest comes from known upper bounds on the minimum Hamming distance of codes. Using the Plotkin bound

$$\delta \leq \frac{2}{3}(1-R) + o(1)$$

gives $R \leq (1-R)/3 + o(1)$, which is asymptotically $R \leq 1/4$, essentially equivalent to[2] (6). The stronger bound (7) is obtained instead by using the best known bound on δ, which is the linear programming bound of [23] adapted to q-ary codes [24], defined implicitly in δ by the inequality

$$R \leq H_q\left(\frac{1}{q}\left(q-1-(q-2)\delta - 2\sqrt{(q-1)\delta(1-\delta)}\right)\right) \tag{9}$$

with $q = 3$ and

$$H_q(t) = t\log_q(q-1) - t\log_q t - (1-t)\log_q(1-t). \tag{10}$$

The bound on R is found by combining (8) and (9), which means solving (9) for equality with $\delta = 2R$.

III. General Case $q \geq k \geq 3$

Our extension to general $q \geq k \geq 3$ is essentially based on the idea of iterating the technique used for $q = k = 3$. To do this, we will need to consider a generalized notion of distance among tuples of codewords and a generalization of Jamison's bound to multiple coverings. The latter is given by this result of Bruen [25].

Lemma 2 ([25]): Let $\mathcal{H}$ be a multiset of hyperplanes in $\mathbb{F}_q^m$. If no hyperplane in $\mathcal{H}$ contains 0 and each point in $\mathbb{F}_q^m \setminus \{0\}$ is covered by at least t hyperplanes in $\mathcal{H}$, then

$$|\mathcal{H}| \geq (m+t-1)(q-1).$$

We need to introduce the following technical lemma.

Lemma 3: Let C be a linear code of dimension m in $\mathbb{F}_q^d$, let $x_1, x_2, \ldots, x_\ell \in C$ be $\ell \leq q-1 \leq m$ linearly independent codewords which are all pairwise distinct in each coordinate and contain no zeros, and let C' be a subcode of C of dimension $m-\ell$ that intersects trivially, only in the origin, the subspace spanned by the x_i's. For $i = 1, \ldots, d$, set $S_i = \mathbb{F}_q \setminus \{0, x_{1,i}, x_{2,i}, \ldots, x_{\ell,i}\}$.

Assume that, for each $c \in C' \setminus \{0\}$, we have $c_i \in S_i$ for at least t values of i. Then

$$m - \ell \leq \frac{q-\ell-1}{q-1}d - t + 1. \tag{11}$$

[1]The main tool used is essentially some form of Jamison's bound both here and in [18], [19]. The difference is mainly a matter of how this tool can be combined with other ideas in coding theory.

[2]Strictly speaking, to obtain the positive ϵ in (6) we need the fact that the Plotkin bound is not tight at positive rates.

Proof: Let G be the (full rank) $(m-\ell)\times d$ generator matrix of the subcode C' and g_i its i-th column. Consider the hyperplanes

$$H_{i,b}=\{v\in\mathbb{F}_q^{m-\ell}\mid v\cdot g_i=b\}\,,\quad i=1,\dots,d\,,\ b\in S_i\,.$$

Since $|S_i|=q-(\ell+1)$, these are $(q-\ell-1)d$ hyperplanes none of which contains the zero vector. On the other hand, by assumption, each $v\in\mathbb{F}_q^{m-\ell}\setminus\{0\}$ is covered at least t times by those hyperplanes. Therefore from Lemma 2 we then have

$$(q-\ell-1)d\geq(m-\ell+t-1)(q-1)$$

which is equivalent to the statement. ∎

We are now ready to state our main result.

Theorem 1: Let C be a linear k-hash code in $\mathbb{F}_q^n$ of rate $R=m/n$ and relative distance δ. Then,

$$R\leq\frac{\delta}{\sum_{i=1}^{k-2}\frac{(q-1)^{\underline{i}}}{(q-2)^{\underline{i}}}}+o(1)\tag{12}$$

where $(q-2)^{\underline{i}}=(q-2)(q-3)\cdots(q-i-1)$.

Proof: We will show that if the rate exceeds the claimed bound we can find, by means of an iterative process, a collection of k codewords $\{0,x_1\dots,x_{k-1}\}$ which do not satisfy the k-hash property. Figure 1 gives a graphical representation of the properties we will require for the codewords. We start with a codeword x_1 of minimum weight $d=\delta_1 n=\delta n$, where we assume without loss of generality that x_1 is non-zero in the first d coordinates. Any set of k codewords in C which includes 0 ad x_1 cannot satisfy the k-hash property in the last $n-d$ coordinates, so we can focus on the first d coordinates and consider the punctured code, call it $C_{[d]}$. Note that puncturing is injective. Indeed, if two distinct codewords $y,y'\in C$ are equal in $[d]$, then the codewords $0,x_1,y-y'$ are all distinct and are not 3-hashed. Hence, the code is not a k-hash code for any $k\geq 3$. This means that i) $C_{[d]}$ is also an m-dimensional subspace in $\mathbb{F}_q^d$ and ii) we can refer to *codewords* without ambiguity as to whether we mean in C or $C_{[d]}$.

We now want to select a codeword x_2 which is linearly independent of x_1 and matches either with 0 or with x_1 in many coordinates. Consider thus the linear subspace of $C_{[d]}$ of dimension m. We now use Lemma 3. Take t which contradicts (11) with $\ell=1$, that is such that

$$\frac{m-1}{n}>\frac{q-2}{q-1}\cdot\frac{d}{n}-\frac{t}{n}+\frac{1}{n}\,.$$

Setting $\delta_2=t/n$, this means taking $\delta_2\in[0,1]$ such that

$$R>\frac{q-2}{q-1}\delta_1-\delta_2+o(1)\,.$$

Lemma 3 then implies that there is a codeword x_2 linearly independent of x_1 such that $x_{2,i}\notin\{0,x_{1,i}\}$ for less than $\delta_2 n$ coordinates in $[1,\dots,\delta_1 n]$. Assume without loss of generality that these are the first coordinates. Then, the three codewords $\{0,x_1,x_2\}$ are not 3-hashed in any of the last $(1-\delta_2)n$ coordinates. This means that any k codewords which include $0,x_1,x_2$ must be k-hashed in one of the first $\delta_2 n$ coordinates. Again we can restrict our attention on the punctured code

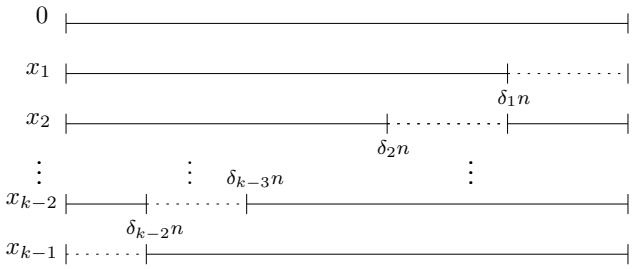


Fig. 1. Codewords used in the proof of Theorem 1. Each codeword collides with one of the previous codewords in each coordinate of the dotted part.

$C_{[\delta_2 n]}$. Note that again the puncturing is injective, for the same reason already mentioned in the first iteration; if distinct codewords y and y' coincide over $[\delta_2 n]$, then $0,\ x_1,\ x_2,\ y-y'$ are 4 distinct codewords which are not 4-hashed (note that x_1 and x_2 cannot equal $y-y'$ since they are non-zero in $[\delta_2 n]$ while $y-y'$ is zero there). Thus $C_{[\delta_2 n]}$ has dimension m. Furthermore, x_1 and x_2 remain linearly independent also when restricted to the coordinates $[\delta_2 n]$. Indeed, if we assume by contradiction that in $[\delta_2 n]$ we have $x_2=\alpha x_1$ for some $\alpha\in\mathbb{F}_q\setminus\{0\}$, then the codeword in $C_{[\delta_1 n]}$ defined by $y=\alpha x_1$ coincides with x_2 in $[\delta_2 n]$ and so, by injectivity of our code restriction to $[\delta_2 n]$, necessarily $y=x_2$, which is impossible since x_2 and x_1 were chosen to be linearly independent in $[\delta_1 n]$. Therefore, since the subspace $C_{[\delta_2 n]}$ has dimension m, we can iterate our procedure invoking Lemma 3 with $\ell=2$.

Continuing this way, at iteration j we have j linearly independent codewords $x_1,\dots,x_j$ in $C_{[\delta_j n]}$ such that

$$|\{0,x_{1,i},\dots,x_{j,i}\}|\leq j\text{ for all }i>\delta_j n\,,$$

and we find a $(j+1)$-th linearly independent codeword over the coordinates $[\delta_j n]$ which is $(j+1)$-hashed with the previous j codewords and 0 only in the first $\delta_{j+1}n$ coordinates, where δ_{j+1} is chosen to satisfy

$$R>\frac{q-j-1}{q-1}\delta_j-\delta_{j+1}+o(1)\,,$$

that is

$$\delta_{j+1}>\frac{q-j-1}{q-1}\delta_j-R.\tag{13}$$

The restriction of the code to the first $\delta_{j+1}n$ coordinates is again injective and thus also preserves the linear independence of $x_1,\dots,x_{j+1}$, because if one of those codewords was a linear combination of the other ones over $[\delta_{j+1}n]$ then it would coincide with the same linear combination taken over $[\delta_j n]$ which contradicts the linear independence of $x_1,\dots,x_{j+1}$ over $[\delta_j n]$.

We iterate this for $j=1,\dots,k-3$, finding $x_1,\dots,x_{k-2}$ such that

$$|\{0,x_{1,i},\dots,x_{k-2,i}\}|\leq k-2\text{ for all }i>\delta_{k-2}n$$

while

$$|\{0,x_{1,i},\dots,x_{k-2,i}\}|=k-1\text{ for all }i\leq\delta_{k-2}n.$$

At this point we can find one last, linearly independent codeword x_{k-1}, such that

$$x_{k-1,i} \in \{0, x_{1,i}, \ldots, x_{k-2,i}\}\,, \text{ for all } i \leq \delta_{k-2} n$$

if (13) is satisfied for $j = k-2$ with $\delta_{k-1} = 0$, that is if

$$R > \frac{q-k+1}{q-1}\delta_{k-2}\,. \tag{14}$$

This gives us $k-1$ codewords $x_1, \ldots, x_{k-1}$ such that the k codewords $\{0, x_1, \ldots, x_{k-1}\}$ are not k-hashed in any coordinates. The condition on R can be obtained by using recursively equation (13) in (14) with initialization $\delta_1 = \delta$. This leads to

$$\frac{q-1}{q-k+1}R > \frac{(q-2)^{\underline{k-3}}}{(q-1)^{k-3}}\delta - R\sum_{j=0}^{k-4}\frac{(q-k+j+1)^{\underline{j}}}{(q-1)^j}$$

which, after rearrangements of the terms, is equivalent to R violating (12). ∎

IV. New upper bounds

In this section, we provide, using the result obtained in Theorem 1, new upper bounds on the rate of linear k-hash codes in $\mathbb{F}_q^n$ for every $q \geq k \geq 3$.

Using the Plotkin bound for $q \geq 3$ we obtain the following corollary.

Corollary 1: Let C be a linear k-hash code in $\mathbb{F}_q^n$ of rate R. Then,

$$R \leq \left(1 + \frac{q}{q-1}\sum_{i=1}^{k-2}\frac{(q-1)^i}{(q-2)^{\underline{i}}}\right)^{-1} + o(1)\,. \tag{15}$$

Proof: By the Plotkin bound we have that a code of length n with relative minimum distance δ and rate R satisfies, for n large enough, the inequality $R \leq 1 - \frac{q}{q-1}\delta$ which implies that

$$\delta \leq \frac{q-1}{q}(1-R)\,. \tag{16}$$

Now, using the upper bound on δ of equation (16) and the bound for linear k-hash codes given in Theorem 1 we obtain

$$R \leq \frac{q-1}{q}\frac{(1-R)}{\sum_{i=1}^{k-2}\frac{(q-1)^i}{(q-2)^{\underline{i}}}} + o(1)\,.$$

Therefore, rearranging the terms we obtain the statement of the corollary. ∎

As done for the case $q = k = 3$ in Section II, we can use the first linear programming bound of [24] to obtain the following corollary.

Corollary 2: Let C be a linear k-hash code in $\mathbb{F}_q^n$ of rate R. Then,

$$R \leq \frac{\delta^*}{\sum_{i=1}^{k-2}\frac{(q-1)^i}{(q-2)^{\underline{i}}}} + o(1)\,,$$

where δ^* is the unique root of the following equation in x

$$\frac{x}{\sum_{i=1}^{k-2}\frac{(q-1)^i}{(q-2)^{\underline{i}}}} = H_q\left(\frac{1}{q}\left(q-1-(q-2)x - 2\sqrt{(q-1)x(1-x)}\right)\right)\,,$$

where $0 \leq x \leq \frac{q-1}{q}$ and H_q is the function defined in equation (10).

TABLE I
Upper bounds on the rate of linear 3-hash codes in $\mathbb{F}_q^n$ for a prime power $q \in [3, 64]$. All numbers are rounded upwards.

q	Corollary 1	Corollary 2	Equation (4)
3	$1/4 = 0.25$	0.2198	0.3691
4	$1/3 = 0.\overline{3}$	0.3000	$1/2 = 0.5$
5	$3/8 = 0.375$	0.3441	0.5694
7	$5/12 = 0.41\overline{6}$	0.3928	0.6438
8	$3/7 = 0.\overline{428571}$	0.4080	$2/3 = 0.\overline{6}$
9	$7/16 = 0.4375$	0.4200	0.6846
11	$9/20 = 0.45$	0.4373	0.7110
13	$11/24 = 0.458\overline{3}$	0.4497	0.7298
16	$7/15 = 0.4\overline{6}$	0.4628	$3/4 = 0.75$
17	$15/32 = 0.46875$	0.4663	0.7554
19	$17/36 = 0.47\overline{2}$	0.4721	0.7646
23	$21/44 = 0.477\overline{27}$	0.4811	0.7790
25	$23/48 = 0.4791\overline{6}$	0.4846	0.7847
27	$25/52 = 0.48\overline{076923}$	0.4877	0.7897
29	$27/56 = 0.482\overline{142857}$	0.4903	0.7942
31	$29/60 = 0.48\overline{3}$	0.4927	0.7982
32	$15/31 \approx 0.483871$	0.4938	$4/5 = 0.8$
37	$35/72 = 0.486\overline{1}$	0.4984	0.8081
41	$39/80 = 0.4875$	0.5013	0.8134
...	...	...	...
64	$31/63 = 0.\overline{492063}$	0.5119	$5/6 = 0.8\overline{3}$

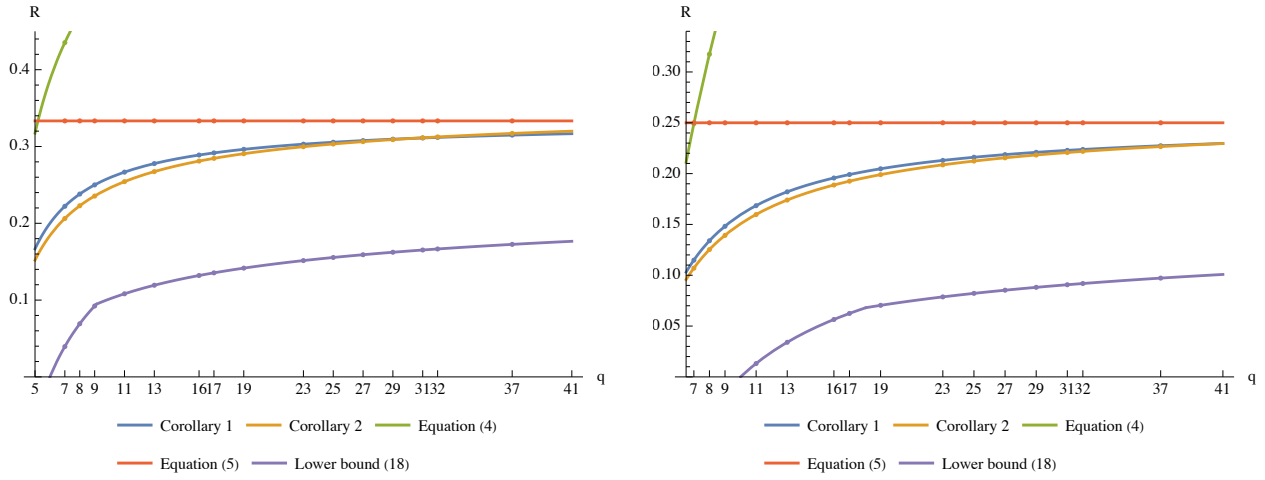
In Table I, we compare the bounds provided in Corollaries 1, 2 and the one given in (4) for $q \in [3, 64]$. It can be seen that the linear programming bound performs better for $q \leq 19$ while for $q \geq 23$ the Plotkin bound gives a better result.

Remark 1: We observe that one could use the second linear programming bound or the straight-line bounds given in [24], [26] to improve the results of Corollaries 1 and 2 for different values of q and k. Here we avoid to show those improvements to keep a simpler presentation of our results.

For a fixed value of k and $q \to \infty$, both bounds given in Corollaries 1 and 2 converge to $1/(k-1)$, which is the same upper bound on the rate of (q,k)-hash codes (not necessarily linear) that one can derive from equation (5). Corollary 1 approaches $1/(k-1)$ from below since the rhs of equation (15) is strictly increasing in q and since $\sum_{i=1}^{k-2}\frac{(q-1)^i}{(q-2)^{\underline{i}}} \geq k-2$, while Corollary 2 does not, see for example Table I where for $k = 3$ and $q = 41, \ldots, 64$ we have upper bounds that exceed $1/2$.

We can compare the bound of Corollary 1 with the one given in equation (4) to obtain the following theorem.

Theorem 2: For every $q \geq k^2$ and $k \geq 4$, the bound of Corollary 1 improves the one of Körner and Marton for general codes given in equation (4).



(a) (b)

Fig. 2. (a) – Comparison between upper and lower bounds for $q \geq 5$ and $k = 4$; (b) – Comparison between upper and lower bounds for $q \geq 7$ and $k = 5$.

Proof: We need to show that

$$\left(1 + \frac{q}{q-1}\sum_{i=1}^{k-2}\frac{(q-1)^i}{(q-2)^{\underline{i}}}\right)^{-1} < \min_{0\leq j\leq k-2}\frac{q^{j+1}}{q^{j+1}}\log_q\left(\frac{q-j}{k-j-1}\right). \quad (17)$$

We lower bound the rhs of equation (17) as follows

$$\begin{aligned}\min_{0\leq j\leq k-2}\frac{q^{j+1}}{q^{j+1}}\log_q\left(\frac{q-j}{k-j-1}\right) &\geq \frac{q^{k-1}}{q^{k-1}}\log_q\left(\frac{q}{k-1}\right)\\ &\geq \frac{1}{2}\left(\frac{q-k+2}{q}\right)^{k-2},\end{aligned}$$

since the function $\log_x(\alpha x)$ is increasing in x for $x \geq 2$ and $0 < \alpha < 1$ and since $k \geq 4$ and $q \geq k^2$. Then, by Bernoulli's inequality we have that

$$\frac{1}{2}\left(\frac{q-k+2}{q}\right)^{k-2} \geq \frac{1}{2}\left(1 - \frac{(k-2)^2}{q}\right).$$

Since the lhs of (17) is less than $1/(k-1)$, in order to prove the statement of the theorem we just need to show that

$$\frac{1}{k-1} \leq \frac{1}{2}\left(1 - \frac{(k-2)^2}{q}\right),$$

but this inequality is satisfied for $q \geq k^2$ and $k \geq 4$. ∎

We conjecture that Theorem 2 still holds also if we relax the hypothesis to $q \geq 2k-3$ and $k \geq 3$. This would imply that, for all the interesting values of q and k (since the asymptotic rate of linear k-hash codes in $\mathbb{F}_q^n$ for $q \leq 2k-4$ is zero), our bound provides the best result.

In support of our conjecture, Table I provides an instance for $k = 3$ where for every $q \geq 3$ the bound of Corollary 1 improves the one of equation (4) and Figures 2a and 2b report the comparison between our bounds and the best known bounds in the literature for $k = 4, 5$ and $q \geq 2k-3$. In addition, we have numerically verified the conjecture for every $k \in [3, 100]$ and $q \geq 2k-3$.

The authors in [7], using classical random coding techniques, provide the following lower bound on the rate of linear k-hash codes in $\mathbb{F}_q^n$ for $q \geq \binom{k}{2}$:

$$R \geq \min\left\{-\frac{1}{k-1}\log_q\left(1 - \frac{q^{\underline{k}}}{q^k}\right), \frac{1}{k-2}\left(1 - \log_q\binom{k}{2}\right)\right\} + o(1), \quad (18)$$

where it can be seen that for q sufficiently larger than k the minimum of (18) is achieved by the first term. This implies that for such values of q and k, the lower bound coincides with the one for general codes given in equation (3). However, there is still a large gap between upper and lower bounds.

In Figures 2a and 2b, we compare our upper bounds and the lower bound of equation (18) for $k = 4, 5$ and $q \geq 2k-3$. We note that both the upper bounds of Corollaries 1, 2 and the lower bound (18) are asymptotically equal to $1/(k-1)$ for a fixed value of k as $q \to \infty$.

ACKNOWLEDGEMENTS

The authors would like to thank Lakshmi Prasad Natarajan for pointing out an error in the original derivation of the main result of this paper. Following his comments, the proof is now also simpler, and the result slightly stronger. The authors would also like to thank Simone Costa for useful discussions on this topic.

REFERENCES

[1] M. Fredman and J. Komlós, "On the size of separating systems and perfect hash functions," *SIAM J. Alg. Disc. Meth.*, vol. 5, pp. 61–68, 1984.

[2] J. Körner and K. Marton, "New bounds for perfect hashing via information theory," *European Journal of Combinatorics*, vol. 9, pp. 523–530, 1988.

[3] E. Arikan, "An upper bound on the zero-error list-coding capacity," *IEEE Trans. Information Theory*, vol. 40, no. 4, pp. 1237–1240, 1994.

[4] S. R. Blackburn and P. R. Wild, "Optimal linear perfect hash families," *Journal of Combinatorial Theory, Series A*, vol. 83, no. 2, pp. 233–250, 1998.

[5] S. Della Fiore, S. Costa, and M. Dalai, "Improved bounds for (b, k)-hashing," *IEEE Transactions on Information Theory*, vol. 68, no. 8, pp. 4983–4997, 2022.

[6] S. Bhandari and J. Radhakrishnan, "Bounds on the zero-error list-decoding capacity of the q/(q – 1) channel," *IEEE Transactions on Information Theory*, vol. 68, no. 1, pp. 238–247, 2022.

[7] L. Bassalygo, M. Burmester, A. Dyachkov, and G. Kabatianski, "Hash codes," in *Proceedings of IEEE International Symposium on Information Theory*, 1997.

[8] S. Della Fiore, A. Gnutti, and S. Polak, "The maximum cardinality of trifferent codes with lengths 5 and 6," *Examples and Counterexamples*, vol. 2, p. 100051, 2022.

[9] S. Kurz, "Trifferent codes with small lengths," *Examples and Counterexamples*, vol. 5, p. 100139, 2024.

[10] S. Bhandari and A. Khetan, "Improved upper bound for the size of a trifferent code," *arXiv preprint arXiv:2402.02390*, 2024.

[11] E. Arikan, "An improved graph-entropy bound for perfect hashing," in *Proceedings of 1994 IEEE International Symposium on Information Theory*, 1994, pp. 314–.

[12] M. Dalai, V. Guruswami, and J. Radhakrishnan, "An improved bound on the zero-error list-decoding capacity of the 4/3 channel," *IEEE Transactions on Information Theory*, vol. 66, no. 2, pp. 749–756, 2019.

[13] S. Costa and M. Dalai, "New bounds for perfect k-hashing," *Discrete Applied Mathematics*, vol. 289, pp. 374–382, 2021.

[14] V. Guruswami and A. Riazanov, "Beating fredman-komlós for perfect k-hashing," *Journal of Combinatorial Theory, Series A*, vol. 188, p. 105580, 2022.

[15] S. Della Fiore, S. Costa, and M. Dalai, "New upper bounds for (b, k)-hashing," in *2021 IEEE International Symposium on Information Theory (ISIT)*. IEEE, 2021, pp. 256–261.

[16] C. Xing and C. Yuan, "Beating the probabilistic lower bound on q-perfect hashing," *Combinatorica*, pp. 1–20, 2023.

[17] S. Costa and M. Dalai, "A gap in the slice rank of k-tensors," *Journal of Combinatorial Theory, Series A*, vol. 177, p. 105335, 2021.

[18] C. Pohoata and D. Zakharov, "On the trifference problem for linear codes," *IEEE Transactions on Information Theory*, vol. 68, no. 11, pp. 7096–7099, 2022.

[19] A. Bishnoi, J. D'haeseleer, D. Gijswijt, and A. Potukuchi, "Blocking sets, minimal codes and trifferent codes," 2023.

[20] S.-L. Ng and P. R. Wild, "On k-arcs covering a line in finite projective planes," *ARS COMBINATORIA-WATERLOO THEN WINNIPEG-*, vol. 58, pp. 289–300, 2001.

[21] A. Calderbank, P. Frankl, R. Graham, W. Li, and L. Shepp, "The sperner capacity of linear and nonlinear codes for the cyclic triangle," *Journal of Algebraic Combinatorics: An International Journal*, vol. 2, no. 1, pp. 31–48, Mar. 1993.

[22] R. E. Jamison, "Covering finite fields with cosets of subspaces," *Journal of Combinatorial Theory, Series A*, vol. 22, no. 3, pp. 253–266, 1977.

[23] R. McEliece, E. Rodemich, H. Rumsey, and L. Welch, "New upper bounds on the rate of a code via the Delsarte-MacWilliams inequalities," *Information Theory, IEEE Transactions on*, vol. 23, no. 2, pp. 157 – 166, mar 1977.

[24] M. Aaltonen, "A new upper bound on nonbinary block codes," *Discrete Mathematics*, vol. 83, no. 2, pp. 139–160, 1990.

[25] A. Bruen, "Polynomial multiplicities over finite fields and intersection sets," *Journal of Combinatorial Theory, Series A*, vol. 60, no. 1, pp. 19–33, 1992. [Online]. Available: https://www.sciencedirect.com/science/article/pii/009731659290035S

[26] T. Laihonen and S. Litsyn, "On upper bounds for minimum distance and covering radius of non-binary codes," *Designs, Codes and Cryptography*, vol. 14, pp. 71–80, 1998.